Fault – Tolerant Design of the IBM POWER6™ Microprocessor

Kevin Reick, Pia N. Sanda, Scott Swaney, Jeffrey W. Kellington, Michael Floyd , Daniel Henderson
IBM Systems Group, Austin, TX
IBM Systems Group, Poughkeepsie, NY

Outline

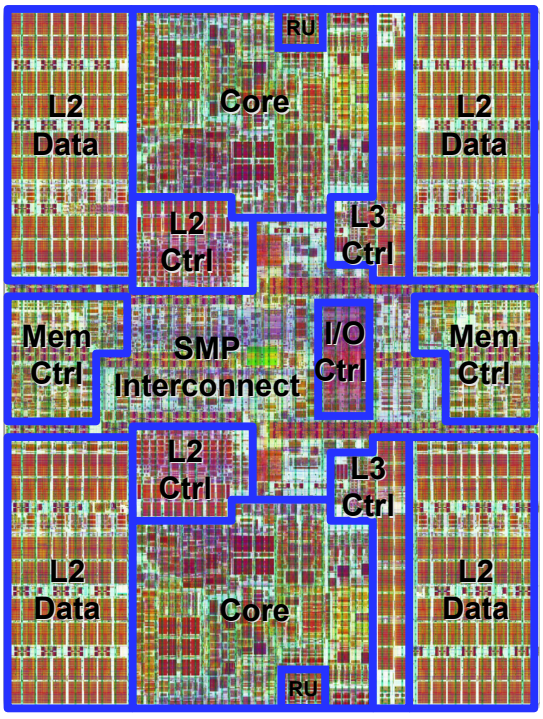
- POWER6™ Overview
- RAS Goals
- New RAS Features
- Detailed description of new features
- Validation of resilience with proton beam accelerated testing
- Conclusions

POWER6 Core

- POWER6 offers ~2X the frequency of POWER5
- POWER6 extends functionality of POWER5 Core
  - Enhanced 2-way SMT with 7 instruction dispatch
  - 64K, 4-way I Cache; 64K, 8-way D Cache
  - Speculative load look-ahead and enhanced data prefetch
  - 2 FXU, 2 FPU, 2 LSU, 1 Branch Unit
  - VMX Unit
  - Decimal Floating Point Unit

POWER6 Chip Overview

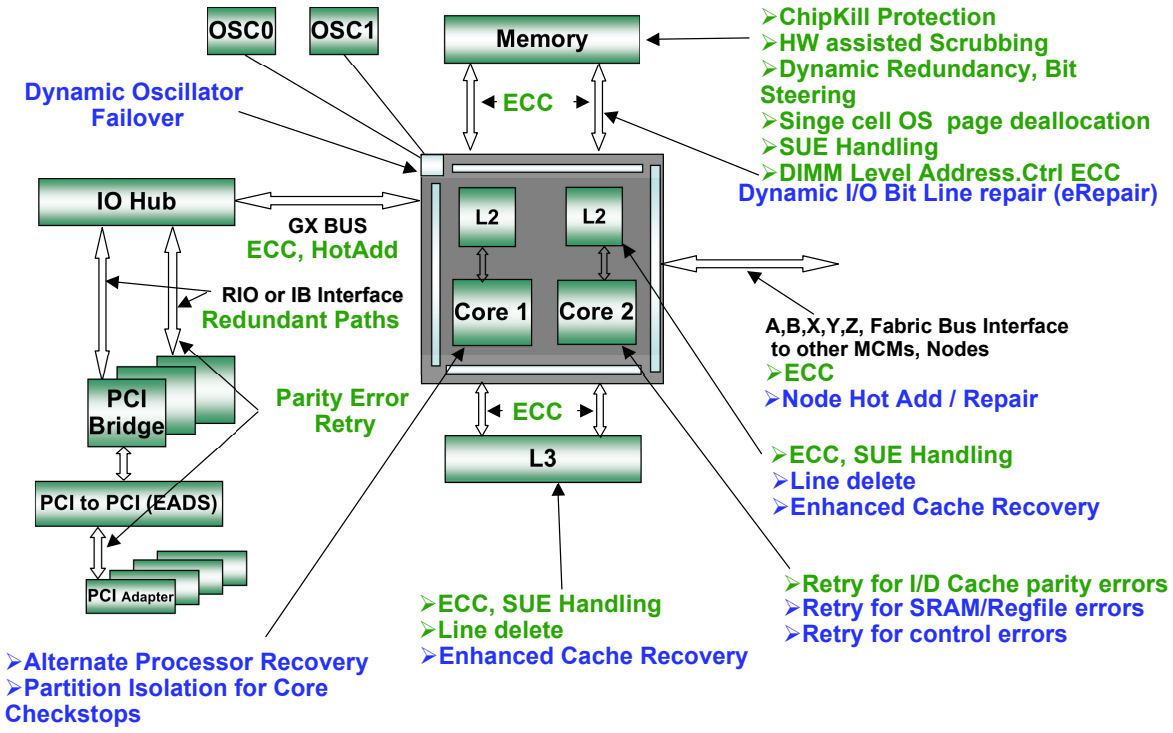
- Ultra-high frequency dual-core chip
  - 7-way superscalar, 2-way SMT core
  - 9 execution units
    - 2LS, 2FP, 2FX, 1BR, 1VMX,1DFU
  - 790M transistors
  - Up to 64-core SMP systems
  - 2x4MB on-chip L2
  - 32MB On-chip L3 directory and controller
  - Two memory controllers on-chip
  - Recovery Unit
- Technology
  - CMOS 65nm lithography, SOI
- High-speed elastic bus interface at 2:1 freq
  - I/Os: 1953 signal, 5399 Power/Gnd



POWER6 RAS GOAL

- Improve upon prior system designs
  - Technology
  - Server consolidation
  - Increasing Customer expectations

Reliability and Availability Features New to POWER6



POWER6 RAS EXECUTION

- Error Detection and Recovery requirements were specified during the High Level Design phase
- Firmware Recovery assists specified early
- The POWER6 RAS design was a collaboration between the System p and System z Processor design teams
- POWER6 shares design methodologies and macros with the System z processor
- Many of the recovery techniques used in POWER6 were initially developed for the System z processor
  - Instruction Retry
  - Alternate Processor Recovery
  - Core checkstop isolation

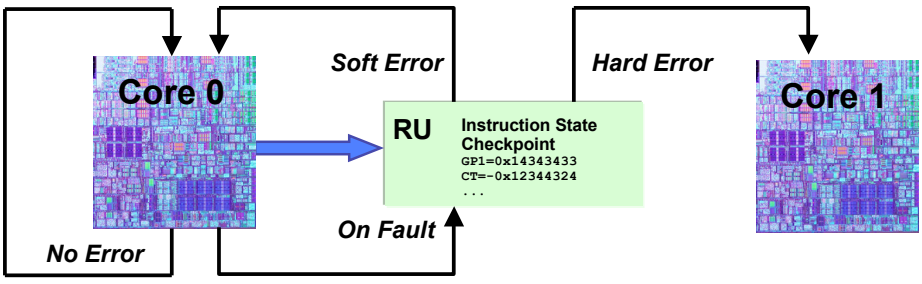
Functions to protect against Core errors

- Processor Instruction retry
  - Retries instructions that had faults
  - Protects against intermittent/soft errors
- Alternate Processor Recovery
  - If instruction retry encounters a second occurrence of the error. (i.e., Solid defect)
  - Moves workload over to an alternate/spare processor
- Processor contained checkstops
  - Limits impact of many processor logic/cmd/ctrl errors to just the processor executing the instruction

Error Detection is first step to Recovery

- 100% ECC protection for caches and interfaces
- >99% of small SRAMs and Register files parity protected
- Dataflow protection
- Protocol checking between functional units
- Control logic protected by parity and consistency checking
- Floating Point Residue Checking
- Queue management (Underflow/Overflow)
- Architected Registers
- Store Data

Core recovery



Non Error Case

- Core architected state is check pointed at every Instruction completion
- Circuitry Checked every cycle

Soft Error Case

- Core restarts from last check point

Hard Error Case

- Hypervisor moves workload to an alternate core

Core Checkstop

- High levels of error detection and isolation were specified early in the design cycle
- Core checkstops fall into two categories:

  Recoverable
  - Core Sparing moves the work to a spare processor

  Non Recoverable
  - Partitions running on the core at the time of the fault are terminated
  - Other partitions are not affected
  - Policy is set by the Hypervisor

Enhanced Cache Recovery

Single bit errors
- Soft errors are purged from the cache to force a refresh of the cell
- Hard errors will result in line delete. Reduces the risk of a double bit error

Multi bit errors
- Hardware will purge and delete the damaged location
- Firmware will dynamically de-configure the core attached to the defective cache

System Recovery of Cache UEs

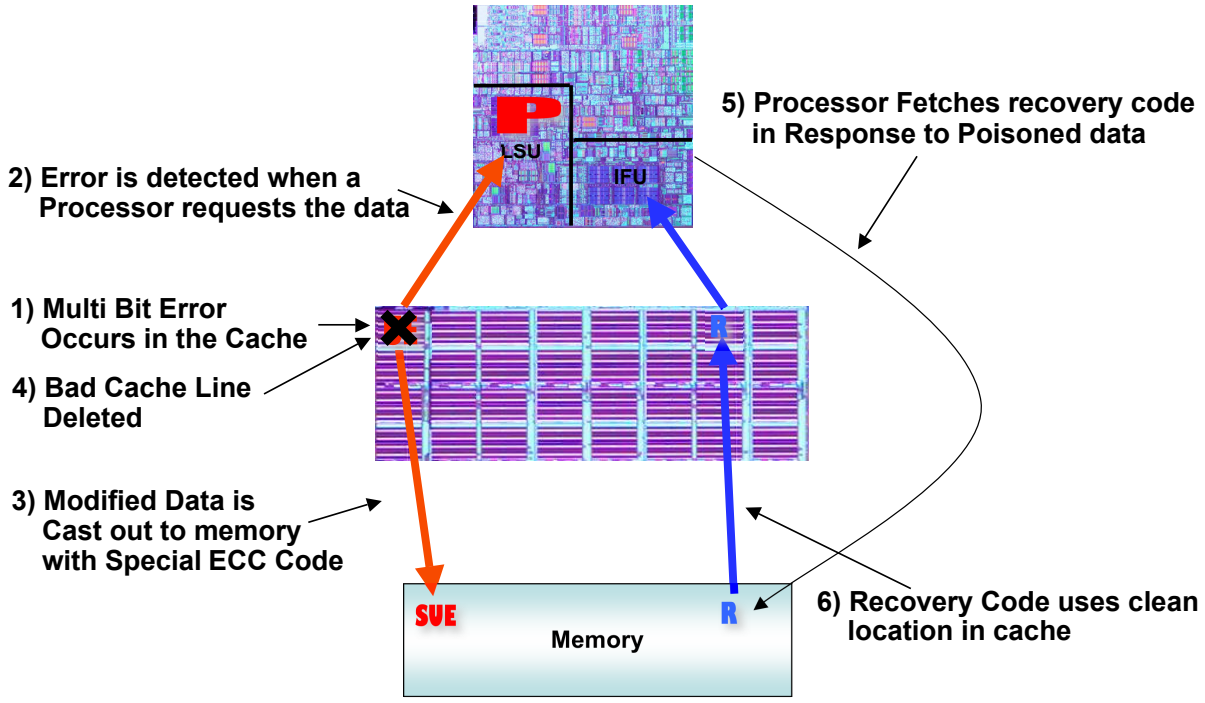
Problem

- POWER6 systems employ System Recovery Code for Uncorrectable errors detected in the Cache Hierarchy
- If cache location is damaged, the same code being used to recover the initial error could be damaged as well

Solution

- POWER6 has automatic purge and delete for L2 and L3 Cache UEs
- Non-modified lines are refetched from Main Store and recovered transparently
- Modified lines are frequently contained to affected application, occasionally resulting in partition outage.

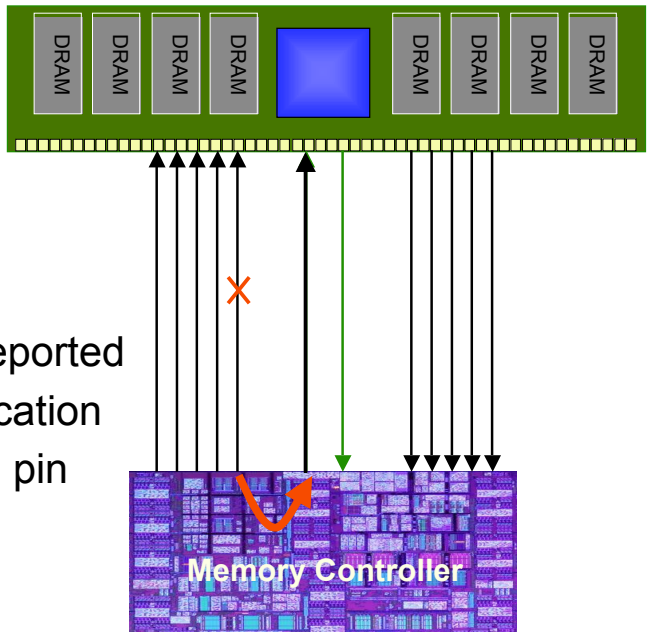
Enhanced Cache Recovery



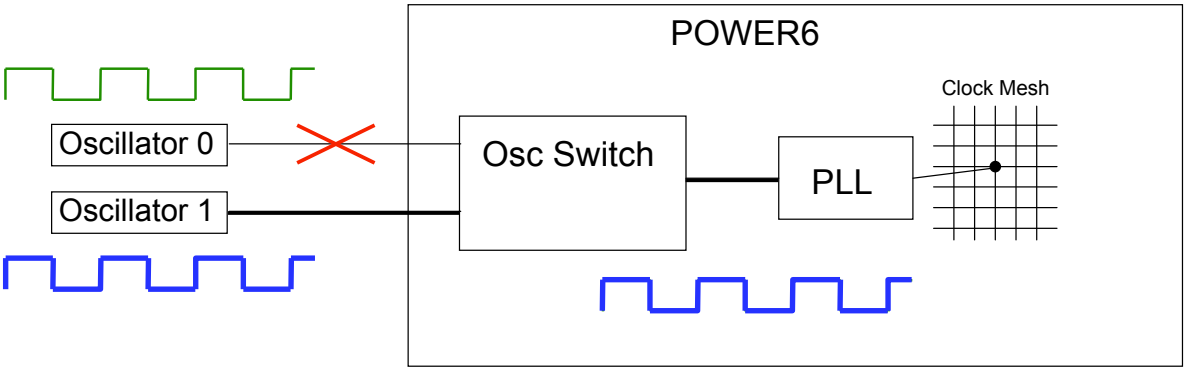
Dynamic I/O Bit Line repair (eRepair)

Memory Data and Control
- Protected with ECC
- Spare Pins added

If a pin breaks
- Correctable errors are reported
- Transparent to the application
- Data redirected to spare pin



Dynamic Oscillator Failover



- System running on Oscillator 0
- Fault Detected on Oscillator 0
- Switch to Oscillator 1 with no disruption to system operation. Eliminates a single point of failure from the system.

Beam Verification of RAS features

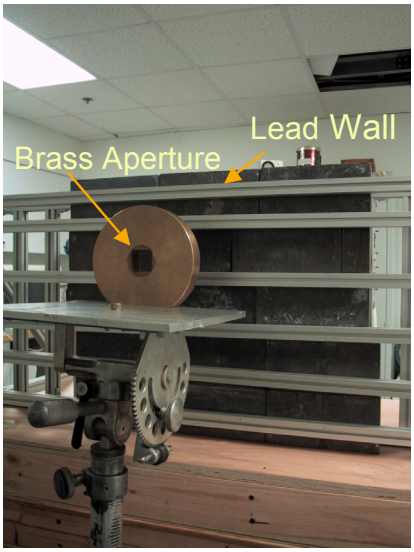
- POWER6 resilience was verified in running systems using two methods of accelerated testing:

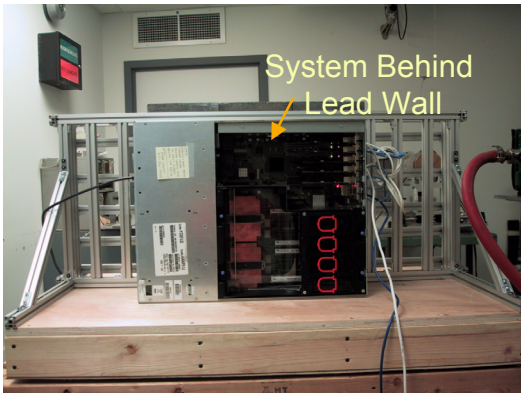
1. Alpha Particle emitters were added to the chip underfill. Used to test Latch and Array resilience
2. Proton Beams were fired though the chip to test the resilience to high energy radiation

POWER6 System in Beamline

- During the Proton Beam experiments 5662 events were recorded.

99.8 % Full Recovery, transparent
0.19 % Resulted in a Partition Outage
0.01 % Resulted in a System Outage





Conclusions

- World Class RAS Depends on:
  - Hardware
    - Error Detection
    - Error Isolation
    - Error Recovery
  - Firmware
    - Error logging and thresholding
  - Hypervisor
    - Intelligent policy decisions for different error scenarios
  - Tightly interlocked design between hardware, firmware and Hypervisor
- Small investment in chip real estate provided resilience to a wide range of soft and hard errors
- Use of fault injection to validate recovery effectiveness proved to be valuable
- POWER6 continues best of breed UNIX Processor and System RAS